Towards a Integrated Evolutionary Model to Study Evolution of Evolution

Charles Rocabert, Carole Knibbe, Guillaume Beslon

Towards a Integrated Evolutionary Model to Study Evolution of Evolution

Charles Rocabert, Carole Knibbe, and Guillaume Beslon

Université de Lyon, CNRS, INRIA, INSA-Lyon, UCB Lyon 1, LIRIS UMR5205,
Beagle Team, F-69622 Lyon, France
charles.rocabert@inria.fr
http://www.evoevo.eu/

Abstract. Variation and Selection are the two core processes of Darwinian Evolution. Yet, both are directly regulated by many processes that are themselves products of evolution. Microorganisms efficiently exploit this ability to dynamically adapt to new conditions. Thus, evolution seems to have optimised their own ability to evolve, as a primary mean to react to environmental changes. We call this process Evolution of Evolution (EvoEvo). In this paper, we propose to use an integrated evolutionary model including complex and evolvable genotype-to-phenotype mapping to study EvoEvo. As exemplified in this paper, the integrated evolutionary model will allow us to decipher the EvoEvo strategies and to offer new hypothesis and predictions on the evolution of microorganisms.

Keywords: Evolution of Evolution, Robustness, Evolvability, Genotype-to-Phenotype Mapping, *In silico* Experimental Evolution

1 Introduction

Life on Earth evolved for billion years in ever changing environments, undergoing smooth or brutal, cyclic or unseen variations. To survive in such conditions, being adapted to the current environment is not enough. Extant organisms had to deal with the evolutionary competition but they also had to deal with the variations of their environment to stay adapted despite the rapid and sometimes profound crisis they had to cope with. How did they do so is an open question. Did the extant organisms survive by chance or did they survive because, being regularly confronted to such crisis, they evolved reaction/adaptation mechanisms?

Experimental evolution, where fast replicating organisms (e.g. bacteria or viruses) are evolved in controlled environments for thousands of generations [3], allows to recover precisely the evolution history of lab strains by reviving frozen samples and performing data analysis. Many evolution results have shown that microorganisms are able to evolve at an amazing speed: in virtually all experimental frameworks that use bacteria or viruses, important phenotypic innovations have emerged in only a few tens of generations [6, 7] and in many cases, evolution tend to be partly reproducible [4, 5]. Microorganisms efficiently use

mutation and selection to dynamically adapt to new conditions. Thus, evolution seems to have optimized their own ability to evolve, as a primary mean to react to environmental changes. We call this process "Evolution of Evolution" (EvoEvo, [8]).

Experimental evolution, despite its explanatory and statistical power, remains a long and costly process. An alternative is to simulate evolution in a computer. However, Evolution of Evolution implies the interaction of a wide range of biological structures (e.g. genome, genetic regulation network, metabolic network, ...), so we need to develop complex models ("The world is complex and we need all the tools that we can muster to understand it" [12]). Following this idea, *in silico* experimental evolution is a growing field in evolutionary biology (reviewed in [23] and [8]), and a lot of theoretical questions have been deciphered with this approach (e.g. [13–17]).

In this paper, we present a new tool to study Evolution of Evolution. we first remind the theoretical background of EvoEvo and why we should use an integrated evolutionary model to study it. In a second part, we will present our model, and finally, our working plan and the perspectives of the integrated evolutionary model.

1.1 What is Evolution of Evolution ?

Variation and Selection are the two core engines of Darwinian Evolution. Yet, both are directly regulated by many processes that are themselves products of evolution (e.g. DNA repair, mutator genes, transposable elements, horizontal transfer, stochasticity of gene expression, sex, network modularity, niche construction...). This ability is the core of the EvoEvo process. Different "Evolution of Evolution" strategies have been proposed in the literature, including regulation of variability, robustness/evolvability strategies, bet-hedging ... However, most of them are poorly characterized and the conditions under which they evolve as well as their consequences are generally unknown.

To develop a better understanding of EvoEvo, we propose to build a conceptual framework based on two important concepts of evolutionary biology: *the genotype-to-phenotype mapping* and the *fitness landscape*. The genotype-to-phenotype mapping summarizes in a single conceptual entity the complex molecular processes by which information flows from the genetic sequence to the organism's phenotype. It thus concatenates in a single abstract process different phenomena such as mRNA transcription, gene translation, protein folding, biochemistry and cell dynamics. The central idea of the fitness landscape is that organisms or populations in evolution can be represented as points on a landscape where the altitude represents the fitness. Selection can be represented by the local gradient of altitude and the mutation can be represented as a random noise added to individual positions.

We propose to study EvoEvo by focusing on four characteristics of the genotype-to-phenotype mapping and the fitness landscape:

Variability. Variability is the ability to generate new phenotypes, by mutations or by stochastic fluctuations. It is a necessary condition for any evolu-

tionary process to take place. However, in biological organisms, the amount of variability is controlled by complex pathways that e.g. correct DNA mismatches or double-strand breaks. Mutational operators are highly diversified, including point mutations, but also large chromosomal rearrangements that can rapidly reshuffle the chromosome organisation, extend or reduce the gene repertoire of an organism or even duplicate its entire genome through whole genome duplication.

Robustness. Although mandatory, variability is a very dangerous process since it permanently produces deleterious mutations that lead to poorly adapted individuals. Robustness may evolve to correct these deleterious effects. It enables evolving systems to support mutational events without loosing fitness through e.g. canalisation or the selection of structures that creates neutral landscapes.

Evolvability. Depending on the genotype-to-phenotype mapping, the proportion of deleterious/neutral/favorable mutational events may change. Evolvability is the ability of a specific genotype-to-phenotype mapping to increase the proportion of favorable events. This can be done by the selection of specific genome structures or by the selection of specific network structures.

Open-endedness. Biological evolution is not directed towards a specific target. On the opposite, evolution has the ability to generate new challenges while evolving by e.g. exploiting new niches created by the evolution of other species.

The central concept of EvoEvo is the following: if the genotype-to-phenotype mapping and the fitness landscape are allowed to change over time, if they can be (indirectly) selected, then they can evolve and acquire properties than could favour evolution in changing environments.

1.2 Why an integrated evolutionary model ?

Computational models have been used to study evolution since the beginning of the 90th [24]. However, since then, most computational models used a partial representation of the genotype-to-phenotype mapping, generally in a fixed, predefined, fitness landscape. By simulating the evolution of such and such organisation level (the genome, the genetic regulation network, the metabolic network, ...), different authors have studied evolution of robustness, evolvability or variability of these specific levels [13, 14, 16, 17]. Yet, EvoEvo is an integrative concept exactly as fitness is. Indeed fitness is the result of the interaction of all the organisation levels of the organism, including its interactions with its environment. Similarly, the robustness/evolvability/variability of the phenotype is the result of the interaction of robustness/evolvability/variability at all the organisation levels of the organism (including its interactions with its environment!). Furthermore, these properties are not independent and they may interact in a cooperative or competitive way (e.g. evolving chaperone proteins reduces the phenotypic variability, thus increasing the robustness). That is why a computational model of EvoEvo must be integrated, including the main organisation levels of the genotype-to-phenotype map (genome, transcription network, metabolic network, phenotype, fitness, population). Moreover, the genotype-to-phenotype mapping must be evolvable, meaning that the complexity and struc-

ture of all these levels must be able to change and that these changes must be able to induce changes of variability/robustness/evolvability. Such a model will necessarily incorporate a large set of parameters and its study is likely to be very difficult. However, it will give rise to new hypothesis and predictions, impossible to obtain with previous models. Indeed, such a model must be seen as a proof-of-concept model [18], used to investigate EvoEvo theory and to test or to generate predictions.

2 Overview of the integrated evolutionary model

In the context of the EvoEvo project, we have designed an integrated model to study Evolution of Evolution. Obviously, such a model cannot include the whole complexity of real organisms. Moreover, one has to keep in mind that our objective is not to study the evolution of such or such organism. Its very aim is to study the evolutionary process and to unravel the EvoEvo strategies that result from a pure Darwininan evolution. That is why we propose to study evolution and EvoEvo in a simplified, abstract, world, designed by an "artificial chemistry" [19]. This artificial chemistry provides a set of objects and a set of rules that govern their interactions. In the model, organisms will be composed of these objects, the rules giving them their dynamic and, ultimately, their fitness.

To build our integrated model, we designed a modular artificial chemistry: the set of objects and the set of rules are split into modules that represent the different organisation levels we want to study as well as the interactions between those levels (e.g. a specific set of rules specifies how the genome is transcribed and translated into proteins). Here we propose to include five levels in our model: the genome, the genetic network, the metabolic network, the fitness and the environment (note that we don't include any "phenotype": the phenotype will simply be the result of the metabolic network dynamic in the organism's environment). These levels are described in the following sections.

Basically, our integrated evolutionary model is an individual based model. Each individual is an asexual virtual cell owning a genome. This genome encodes a regulatory network and a metabolic network. The metabolic network can uptake and convert nutrients from the environment, resulting in the ability (or not) for the cell to divide. Dividing cells form a population that grows on a two dimensional environmental grid providing fresh nutrients, but also nutrients and waste released by cells, actively or after death. This dynamical process results in a modification of the environment and allows for the emergence of complex ecosystems. It also creates the conditions for local competition between cells.

In the model, the molecular structure of the organisms is entirely defined by their genome (possibly with an interaction with the organism's environment). This genome can undergo mutation at each replication (point mutations and large rearrangements). The interaction between this variation process and the competition process described above results in a Darwinian evolution: individuals become more and more adapted to their environment and able to replicate more an more efficiently. However, by doing so, they also modify the shared

environment (e.g. by releasing new metabolites), thus changing their own evolutionary conditions...

In this part, we present our formalism, inspired from [16] and [15], and describe the integrated evolutionary model. In the following section, we will present an example of the evolutionary dynamic observed in the model.

2.1 Description of the genome level

Genome structure. Each organism owns a circular string of functional/non-functional elements. The genome is a coarse-grained genome, inspired on [16], and defined as a list of functional or non-functional elements mathematically defined as n-tuples[1] (the elements of our genome). Each functional tuple is parametrised to define its function and its number of dimensions n (depending on the type). Five types of tuples are defined in our artificial chemistry:

E: Tuples coding for enzymes performing reactions in the metabolic network, via the following Michaelis-Menten equation:

$$\frac{d[p]}{dt} = \frac{k_{cat}.[E].[s]}{k_m + [s]} \tag{1}$$

with $s, p \in \mathbb{N}^*$, $k_{cat} \in \mathbb{R}$ and $k_m \in \mathbb{R}^+$ being encoded in the tuple, $[s]$ and $[p]$ the concentrations of the metabolites s and p, and $[E]$ the enzymatic concentration (we assume that the concentration of free enzymes $[E]$ is always equal to the total concentration $[E_T]$),

TF: Tuples coding for transcription factors. Each transcription factor i owns a binding site identification tag $j \in \mathbb{Z}$ and an affinity A_{ij} for this binding site,

BS: Tuples coding for binding sites specifying which transcription factor may bind to them via their own identification tag $\in \mathbb{Z}$,

P: Tuples coding for promoters determining where the transcription should start. Each promoter i owns a basal expression level β_i,

NC: Non functional tuples constituting the non coding part of the genome.

Binding sites directly flanking a promoter regulate its transcriptional activity. The enhancer site directly precedes the promoter and is made of one or more contiguous binding sites. The operator site directly follows the promoter and is also made of one or more contiguous binding sites. Transcription factors that bind on the enhancer site increase the transcriptional activity. On the opposite, transcription factors that bind on the operator site down-regulate the promoter activity (see figure 1.1). As in R-aevol [15], a promoter has a basal level activity β. TF or E tuples following the operator site are transcribed, thereby allowing for operons. Downstream of the operator site, any tuple other than TF or E makes the transcription stop. To be functional, the promoter can be flanked by binding sites or not, but TF or E tuples must immediately follow the regulation unit (enhancer site + promoter + operator site, see figure 1.2).

[1] A n-tuple is an ordered list $(x_1, x_2, ..., x_n) : T_1 \times T_2 \times ... \times T_n$ with T_i the product type of x_i (e.g. $\mathbb{R}$, $\mathbb{N}$, ...).

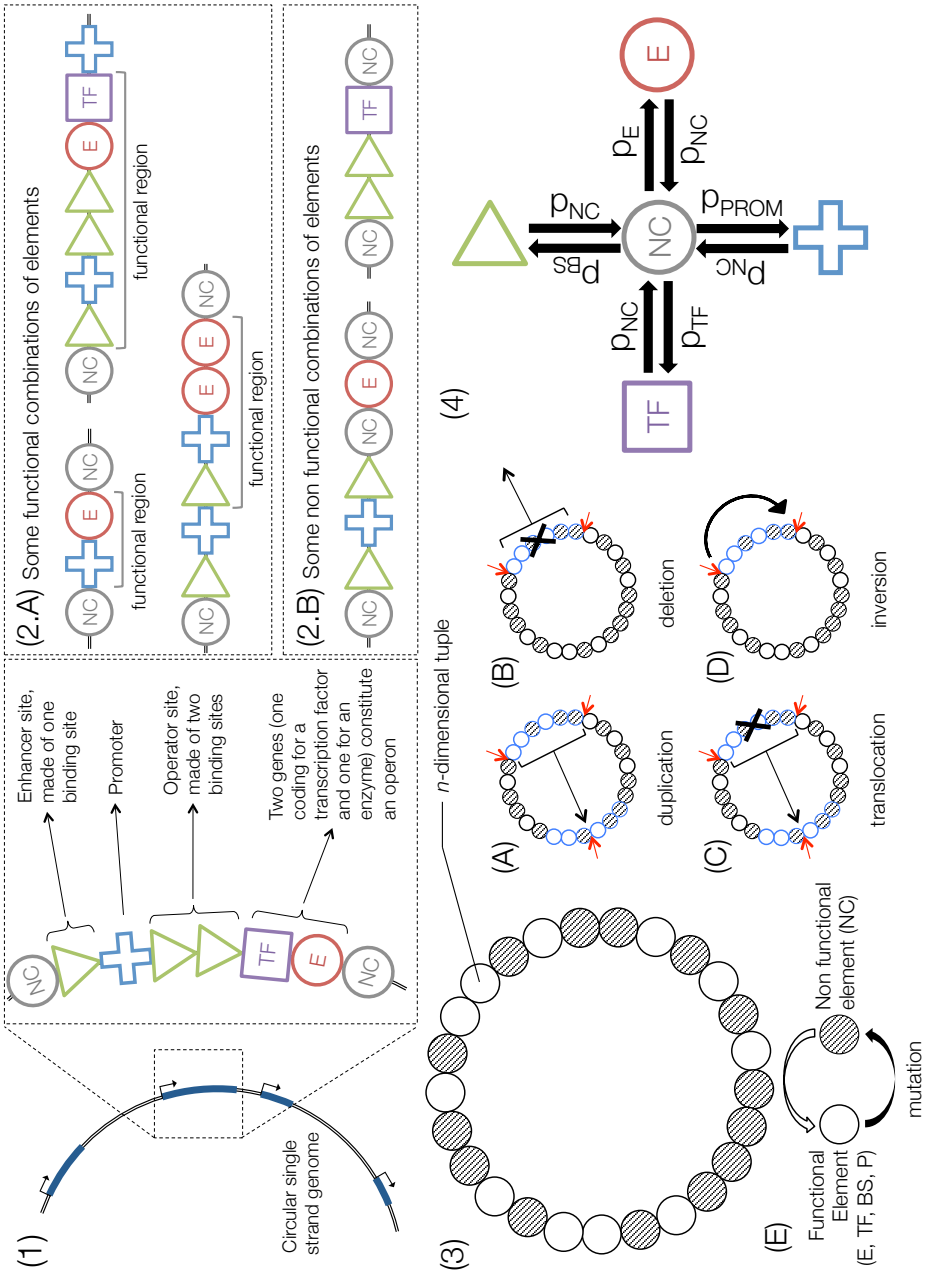

Fig. 1. (1) A functional region starts with a promoter, possibly flanked by an enhancer site and/or an operator site. All contiguous E or TF tuples following the operator site are transcribed. The first tuple of another type interrupts the transcription (here a NC tuple). In this case, the functional region is an operon. **(2.A)** The promoter can be flanked by binding sites or not, but TF or E tuples must immediately follow the regulation unit (enhancer + promoter + operator). **(2.B)** On left, a non coding tuple interrupts the transcription. On right, the promoter is missing. **(3)** The genome is a circular single-strand sequence of tuples. At each replication, the genome undergoes mutations: (A) large duplications, (B) large deletions, (C) translocations, (D) inversions. Red arrows symbolize breakpoints in the sequence. (E) point mutations and breakpoints can unfunctionalise or functionalise a tuple. **(4)** Functional types (E, TF, BS and P) can be unfunctionalised with probability p_{NC}. A NC tuple can be restored to one type or another depending on 4 mutations rates: p_{BS} is the probability to be restored in a binding site (resp. p_{PROM}, p_{E} and p_{TF}).

Mutational operators. The genome undergoes point mutations and large rearrangements during replication. If a tuple undergoes a point mutation, it operates a jump in the tuple space by adding a n-dimensional random vector. A tuple can be unfunctionalised by a point mutation or during a large rearrangement if it is located on a breakpoint. Non coding tuples can also be restored into one or another functional type, however it is impossible to mutate directly from a functional type to another (see figure 1.4). All the mutation rates are configurable. The genome also undergoes large chromosomic rearrangements: duplications, deletions, inversions, and translocations. The various types of mutation can modify existing tuples, but also create new tuples, delete some existing tuples, modify the length of the non coding regions, modify tuple order...

2.2 Description of the genetic regulation network level

The genetic regulation network (GRN) is computed from the interactions of transcription factors (TF) and binding sites (BS) elements. Its activity is computed in four steps:

1. The activity $A_s(t)$ of each binding site s is $A_s(t) = \sum_j c_j(t).A_{js}$ with $c_j(t)$ the concentration of the transcription factor j at time t and A_{js} the affinity of this transcription factor for the binding site s,
2. We then compute the activity of the enhancer site $E_i(t)$ and of the operator site $O_i(t)$ flanking the promoter i:
$$\begin{cases} E_i(t) = \sum_{j \in enhancer_i} A_s(t) \\ O_i(t) = \sum_{j \in operator_i} A_s(t) \end{cases} \tag{2}$$
3. Then, the transcription rate e_i over time of the promoter i is computed, possibly including a stochastic component [20] represented by a transcriptional noise η genetically encoded in the promoter [22]. Note that η mutates as all the elements of the tuple, allowing for evolution of stochastic gene expression [21]. Then the transcription rate e_i of each promoter i is given by an Hill-like function:
$$e_i(t) = \beta_i . \left(\frac{\theta^n}{O_i(t)^n + \theta^n} \right) . \left(1 + \left(\frac{1}{\beta_i} - 1 \right) \left(\frac{E_i(t)^n}{E_i(t)^n + \theta^n} \right) \right) + \xi_i(t) \tag{3}$$
with β_i the basal expression level of the promoter i, n and θ being constant coefficients that determine the shape of the Hill function. $\xi_i(t)$ is a random number drawn from the gaussian distribution $\mathcal{N}(0, \eta_i)$.
4. The transcription rate e_i is applied to each E or TF tuple being controlled by the promoter i, such that each protein product (enzyme or transcription factor) has its own concentration regulated through a synthesis-degradation rule, depending on e_i:
$$\frac{\partial c_i}{\partial t} = e_i(t) - \phi i(t) \tag{4}$$
where ϕ is a temporal scaling constant representing the protein degradation rate.

2.3 Description of the metabolic network level

Enzymes performing reactions in the metabolic network are encoded by tuples of type E. If $s \neq p$, the reaction takes place in the cytoplasm of the cell. If $s = p$, the enzyme is a inflowing or outflowing pump depending on the sign of k_{cat}. For each cell, the whole set of reactions defines an ordinary differential equations (ODE) system, which is solved numerically.

Some metabolic products are essential for the cell's growth, and some other are intermediate products or waste. In the integrated evolutionary model, prime numbers are considered to be essential metabolites: their production contribute to the growth rate by increasing the probability to produce offspring. Over-producing metabolites can also lead to cell's toxicity. Hence, one can define toxicity thresholds for essential and non essential metabolites. Over-reaching a toxicity threshold impairs cell's fitness. Finally, during replication, daughter cells share cytoplasmic content at division (proteins and metabolites). It is also possible to define energy constraints in the artificial chemistry, such that cells must perform catabolic reactions to earn energy and produce essential metabolites.

2.4 Coupling the genetic and the metabolic networks

Bacteria are able to sense their environment by detecting the presence of a particular molecule or signal, and to give an appropriate answer by updating their gene expression profile. In the integrated evolutionary model, co-enzymes can repress or activate transcription factors activity. This is done by adding three elements to the transcription factor tuple (TF): A co-enzyme identification tag $\in \mathbb{N}^*$, a free activity (A_{free}, boolean) and a bound activity (A_{bound}, boolean).

A metabolite $m \in \mathbb{N}^*$ can repress or activate a TF acting as a co-enzyme:

- If $A_{bound} = true$ and $A_{free} = false$, the co-enzyme activates the TF,
- If $A_{free} = true$ and $A_{bound} = false$, the co-enzyme inhibits the TF,
- If $A_{free} = false$ and $A_{bound} = false$, the TF is never active,
- If $A_{free} = true$ and $A_{bound} = true$, the TF is always active.

Finally, the concentration $c_i(t)$ of the TF and the concentration $coE_i(t)$ of the co-enzyme are combined (depending on the values of A_{free} and A_{bound}) to compute the active fraction of $c_i(t)$.

2.5 Description of the population and environment levels

Individuals "live" on a two dimensional grid, each grid site containing at most one individual. The physical environment is described at the grid level: each grid site contains a list of free metabolites, each with its concentration level. Those free metabolites diffuse with a diffusion rate D and are degraded with a degradation rate D_g.

Individuals compete for the free metabolites and to produce offspring in empty sites. Individuals interact with their local environment by pumping metabolites in and out and releasing their content at death. Metabolites can also diffuse

through the cell membrane at rate D_m. In this case, pumps are active mechanisms that the cell can use to maintain an internal concentration different from the external one.

At each simulation time step, organisms are evaluated and either killed, updated or replicated depending on their current state:

1. An active cell can go to death. Death probability follows a Poisson law. At death, cell content is released in the local environment,
2. If the cell do not die and is unable to divide (e.g. because there is no free space in its neighbourhood), its genetic regulation network and metabolic network are updated, and its score is computed from its metabolic concentrations state vector X through the score function F (e.g. the sum of essential metabolite concentrations).
3. For each empty grid site, all active neighbours compete to select the replicating individual, depending on relative fitnesses. To avoid biases, empty grid cells are updated in a random order.

This simple framework allows to model different real experimental setups, including serial plates or chemostat [23]. Similarly, some individuals can be regularly picked up in the environment to seed a new colony, thus mimicking a mutation accumulation experiment.

3 Simulation example

The integrated evolutionary model allows the user to choose the complexity of the simulation by removing or including biological organisation levels. As an example, we present here the results of a simulation where cells evolved their metabolic network in a highly variable environment of size $32*32$ for 100000 time steps (approximatively 10000 generations). Some features have been shut down: the genetic regulation network, the transcriptional noise, energy constraints and membrane permeability. The meaningful parameters of the simulation are displayed in table 1 (among 55 parameters). The score function F is the sum of essential metabolite concentrations produced by the cell, and is updated at each time step of the cell's life. Metabolite $m = 1$ at concentration $[m] = 20.0$ is introduced in the environment as an exogenous nutrient, at random sites with a probability $p = 1e - 05$ per grid site per simulation time step, making the environment a low nutrient and variable one.

3.1 Results

The simulation provides a large set of outputs and statistics, available in a dedicated HTML viewer[2]. Lineage and phylogenic trees are also dynamically computed during the simulation, and displayed in the viewer. Examining the

[2] The viewer of this simulation is available at `http://liris.cnrs.fr/~crocaber/SimulationViewer/viewer/viewer.html`

Table 1. Meaningful parameters used for the simulation detailed in this section, among 55 parameters.

Parameter	Symbol	Value	Parameter	Symbol	Value
Simulation time		1e+05	Stochastic gene expression		NO
Energy constraints		NO	Membrane permeability		NO
Metabolic inheritance		YES	Selection threshold	θ	0.6
Selection pressure	p	10	Grid width	W	32
Point mutation rate		1e-04	Grid height	H	32
Duplication rate		1e-04	Unfunctionalisation rate	p_{NC}	1e-04
Deletion rate		1e-04	Functionalisation rate	p_E	1e-06
Translocation rate		1e-04	Metabolites toxicity threshold		1
Inversion rate		1e-04	Death probability	p_{death}	0.05
Diffusion rate	D	0.01	Degradation rate	D_g	0.001

lineage of the last best individual of the simulation gives a first idea of the evolution of main simulation variables, as exemplified in the figure 2:

Score. The score of each generation is averaged through the cell's lifespan, and is boxed by the minimum and maximum scores reached by the cell during its life. Here, the score tends to increase but is sometimes impaired (at approximatively 2000, 7000 and 10000 generations).

Cytoplasm metabolic content. Each cell inherits an half of the cytoplasm of its mother (light blue), and then starts to accumulate nutrients (dark blue). The score is strongly correlated with the evolution of the cytoplasm content.

Genome size. Thanks to large rearrangements, the genome size varies through generations. Here, the genome undergoes strong size variations. Large genome size duplications seems to be correlated with score impairment. Proportions of functional (coding) and non-functional (non coding) elements indicate that non coding elements represent a large part of the genome.

Genetic redundancy for metabolism. Several genes in the genome can code for the exact same enzyme in the metabolic network. This genetic redundancy associated with the metabolic network is strongly correlated with genome duplications and deletions.

To grow (i.e. increase their fitness), cells need to produce and accumulate essential metabolites. Yet, toxicity thresholds and limited nutrient concentrations in the environment strongly restrict fitness improvement. Thus, the best lineage progressively increases its fitness but sometimes undergoes rapid fitness drops (we call them "crisis"). To understand the origin of these events, we now examine population and environment levels. The population size is variable, alternating plateauing phases not far from the maximum grid capacity (e.g. between 35000 and 50000 time steps), and size falls during crises (e.g. at 60000 or 100000 time steps, see figure 3.1). Evolution of environmental metabolic concentrations shows a rapid diversification (see figure 3.2), with temporary accumulation of one specific essential metabolite at high levels, followed by metabolic sweeps. Sweeping

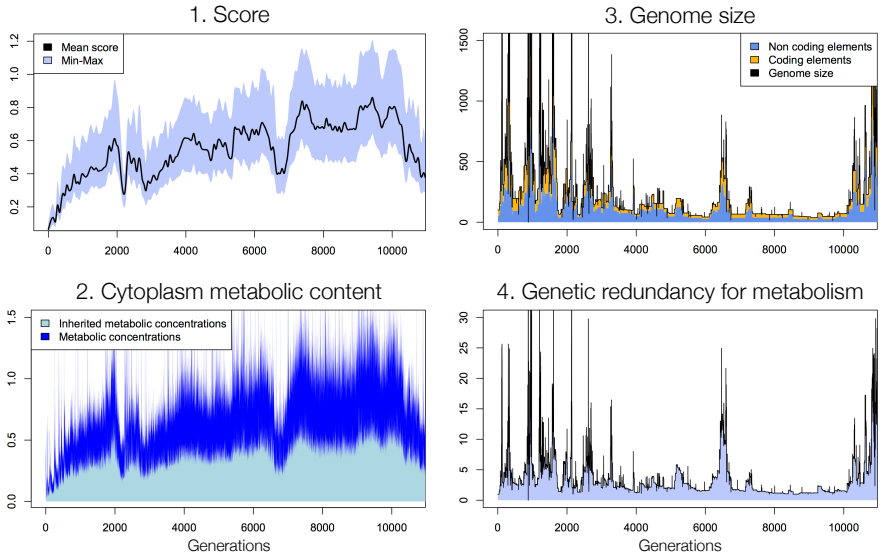

Fig. 2. Some variables showing the evolution of the last best individual lineage. **1.** Evolution of the cell's score. **2.** Evolution of the cytoplasm metabolic content. **3.** Evolution of the genome size. **4.** Evolution of the genetic redundancy associated to the metabolism. The meaning of each variable is explained in the results section 3.1.

periods are associated with population crisis. To investigate more deeply simulation results, the integrated evolutionary model allows us to go back in the evolutionary history by "reviving" ancestor populations saved in backup files. Each 1000 time steps, we computed the trophic network of the population, classifying cells depending on their interaction with other cells and their environment. The cell's state (cytoplasm content, enzymatic reactions, inflowing and outflowing pumps) defines a "profile" used to build a graph, each node corresponding to one specific profile (or group). Edges of the graph represent consuming (solid edges) or helping (dashed edges) relationships. We distinguish three trophic levels:

Level 0: Trophic groups eating and converting nutrients that are not produced by other individuals. These nutrients can be $m = 1$ (the nutrient provided randomly) or any other nutrient released previously by a now extinct colony,

Level 1: Trophic groups eating environment nutrients and the product of other cells,

Level 2: Trophic groups only eating nutrients actively produced by other cells.

We recovered the proportion of groups belonging to each level (figure 3.3) and the numbers of cells per group (figure 3.4). We also recovered the trophic level of the last best individual lineage, as shown at the bottom of the figure 3.4. The three levels undergo strong fluctuations during the simulation. Figure 3.3

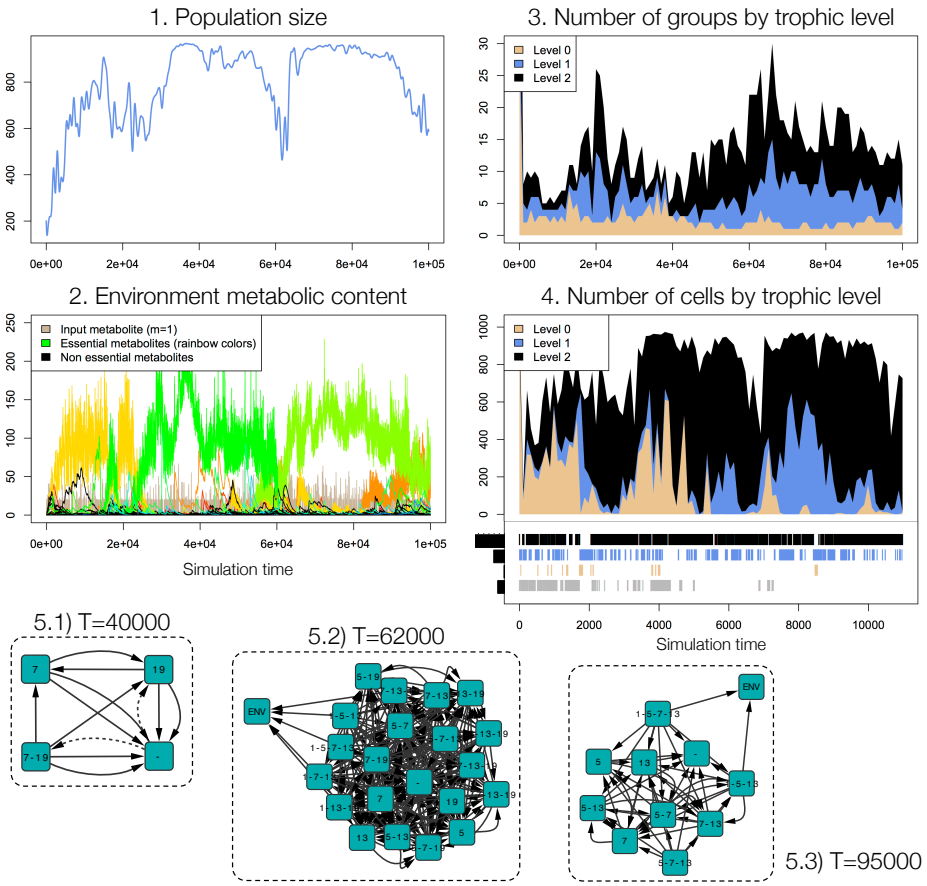

Fig. 3. Evolution of the population and the environment. **1.** Evolution of the population size. **2.** Evolution of metabolic concentrations in the environment. **3.** Evolution of the number of trophic groups per trophic level. **4.** Evolution of the number of cells per trophic level. **5.** Representation of three typical trophic networks. Node labels represent metabolites pumped in by the trophic group (the "uptake profile"). Solid edges represent a consuming activity (e.g. in **5.1**, group 19 consumes group 7 metabolic products). Dashed edges represent "helping" activity, i.e. active release. "-" groups have no inflowing pumps. "ENV" group is the environment, taking only into account the exogenous nutrient.

indicates that the number of groups per level fluctuates, increasing during crisis, and decreasing in between. The number of groups belonging to level 0 tends to be maintained at low levels. Between crisis, the level 2 is the most representative one in the population, but the situation is reverted during crisis, where levels 0 and 1 increase (see figure 3.4). The bottom part of figure 3.4 indicates that trophic groups are not monophyletic. Indeed, the lineage of the last best individual regularly transits between levels, even if level 2 is the most frequent one (grey level represents cells consuming environment wastes, not produced anymore by other cells). Variations in figure 3.3 is supported by trophic network examples. In figure 3.5.1 (at 40000 time steps, during a plateauing phase), interacting groups are only level 2 groups. Level 0 groups do not produce anymore nutrients for those groups ($m = 7$ and $m = 19$), that are thus devoted to extinction. The trophic network 3.5.2 (at 62000 time steps, during a crisis) shows a large complexification of the network, with a new metabolite being produced ($m = 13$). The trophic network 3.5.3 (at 95000 time steps, during a plateauing phase) shows a complexity reduction, but still maintained at a higher level than for the trophic network 3.5.1.

3.2 Discussion

At the beginning of the simulation, cells quickly evolve a metabolic network that uptakes the exogenous nutrient ($m = 1$, in light brown on figure 3.2), and convert it in essential metabolites. The task is tough, the nutrient being rare and localised, so the population is at low level. At death, cells release their cytoplasm content in the environment, thus enriching it. This accumulation of nutrients is followed by the rapid growth of level 2 cells, specialised in consuming the product of others cells. The decline of level 0 and level 1 cells leads to the depletion of level 2 cells nutrient, and the whole population decreases. At some point, new level 0 and level 1 cells convert the exogenous metabolite in a new essential one, cycling this behavior. Some questions arise from those results:

(1) Why don't we observe the fixation of level 1 cells by niche exclusion? The reason could be the cost to convert the exogenous metabolite, such that level 2 cells always grow faster if nutrients are available,

(2) Why do we observe a innovation process at each population crisis, instead of a simple negative frequency-dependent behaviour? If level 0 and level 1 cells compete with level 2 ones for available space, a strategy could be to innovate and produce unseen metabolites that level 2 cells cannot consume. The evolution of concentrations in the environment (figure 3.2) could support this conjecture. Indeed, new essential metabolite production always starts at the very beginning of population crisis,

(3) What is the role of large genome size duplications during crisis ? And what is the link with trophic network complexity variations ? Duplicating the genome size increases enzymatic dosage (enzymatic reactions go faster) but also the mutation rates. If a competition occurs between levels, indirect selection could favour mutators that try to escape their niche to find a new metabolic strategy.

These questions are clearly relevant in an EvoEvo perspective and this preliminary study can be continued by e.g. comparing these results with simulation in which the same amount of food is provided but in an homogeneous way. Similarly, the effect of genomes duplication must be investigated further to understand more precisely their origin and consequences.

4 Perspectives

In this example, we observed the emergence of a complex negative frequency-dependent behaviour, associated with regular metabolic innovations. To understand it, we investigated several organisation levels of the model, linking for example the evolution of the genome structure to the metabolic network and the ecosystem levels. Our preliminary results remind situations observed in bacteria, e.g. the negative frequency-dependent strategy observed in one of the Lenski E. coli lineages [25].

The integrated evolutionary model offers complex evolutionary outcomes, but also the tools to decipher them (regular backups, phylogenies, a large set of statistics, the recovering of trophic networks...). The complexity of the genotype-to-phenotype mapping and of the fitness landscape allows us to investigate the evolutive interactions of a large set of biological organisation levels, and then to deeply explore EvoEvo. Several questions could be investigated, such that the evolution of robustness, evolvability or open-endedness, with the hope to open the door towards a unified theory of Evolution of Evolution.

Acknowledgments This work was supported by the European Commission 7th Framework Programme (FP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427).

References

1. Elena, S. F., Lenski, R. E.: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics, 4(6), 457-469 (2003).
2. Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I., Whitlock, M. C. Experimental evolution. Trends in ecology and evolution, 27(10), 547-560 (2012).
3. Philippe, N., Crozat, E., Lenski, R. E., Schneider, D.: Evolution of global regulatory networks during a long term experiment with Escherichia coli. Bioessays, 29(9), 846-860 (2007).
4. Hekstra, D. R., Leibler, S.: Contingency and statistical laws in replicate microbial closed ecosystems. Cell, 149(5), 1164-1173 (2012).
5. Raeside, C., Gaff, J., Deatherage, D. E., Tenaillon, O., Briska, A. M., Ptashkin, R. N., ... Schneider, D.: Large chromosomal rearrangements during a long-term evolution experiment with Escherichia coli. MBio, 5(5), e01377-14 (2014).
6. Rainey, P. B., Travisano, M.: Adaptive radiation in a heterogeneous environment. Nature, 394(6688), 69-72 (1998).

7. Zhang, Q., Lambert, G., Liao, D., Kim, H., Robin, K., Tung, C. K., ... Austin, R. H.: Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. Science, 333(6050), 1764-1767 (2011).
8. Hindr, T., Knibbe, C., Beslon, G., Schneider, D.: New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nature Reviews Microbiology, 10(5), 352-365 (2012).
9. Tenaillon, O., Taddei, F., Radman, M., Matic, I.: Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. Research in microbiology, 152(1), 11-16 (2001).
10. Kirschner, M., Gerhart, J.: Evolvability. Proceedings of the National Academy of Sciences, 95(15), 8420-8427 (1998).
11. Reisinger, J., Miikkulainen, R.: Selecting for evolvable representations. In Proceedings of the 8th annual conference on Genetic and evolutionary computation (pp. 1297-1304). ACM (2006).
12. Peck, S. L.: Simulation as experiment: a philosophical reassessment for biological modeling. Trends in Ecology and Evolution, 19(10), 530-534 (2004).
13. Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., Adami, C.: Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature, 412(6844), 331-333 (2001).
14. Knibbe, C., Coulon, A., Mazet, O., Fayard, J. M., Beslon, G.: A long-term evolutionary pressure on the amount of noncoding DNA. Molecular biology and evolution, 24(10), 2344-2353 (2007).
15. Beslon, G., Parsons, D. P., Pena, J. M., Rigotti, C., Sanchez-Dehesa, Y.: From digital genetics to knowledge discovery: Perspectives in genetic network understanding. Intelligent Data Analysis, 14(2), 173-191 (2010).
16. Crombach, A., Hogeweg, P.: Evolution of evolvability in gene regulatory networks. PLoS computational biology, 4(7), e1000112 (2008).
17. Cuypers, T. D., Hogeweg, P.: Virtual genomes in flux: an interplay of neutrality and adaptability explains genome expansion and streamlining. Genome biology and evolution, 4(3), 212-229 (2012).
18. Servedio, M. R., Brandvain, Y., Dhole, S., Fitzpatrick, C. L., Goldberg, E. E., Stern, C. A., ... Yeh, D. J.: Not Just a Theory?The Utility of Mathematical Models in Evolutionary Biology. PLoS biology, 12(12), e1002017 (2014).
19. Dittrich, P., Ziegler, J., Banzhaf, W.: Artificial chemistries - a review. Artificial life, 7(3), 225-275 (2001).
20. Elowitz, M. B., Levine, A. J., Siggia, E. D., Swain, P. S.: Stochastic gene expression in a single cell. Science, 297(5584), 1183-1186 (2002).
21. Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., Weissman, J. S.: Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature, 441(7095), 840-846 (2006).
22. Roberts, E., Magis, A., Ortiz, J. O., Baumeister, W., Luthey-Schulten, Z.: Noise contributions in an inducible genetic switch: a whole-cell simulation study. PLoS computational biology, 7(3), e1002010 (2011).
23. Mozhayskiy, V., Tagkopoulos, I.: Microbial evolution in vivo and in silico: methods and applications. Integrative Biology, 5(2), 262-277 (2013).
24. Adami, C.: Digital genetics: unravelling the genetic basis of evolution. Nature Reviews — Genetics, 7, 109 (2006).
25. Rozen, D. E., Philippe, N., Arjan de Visser, J., Lenski, R. E., Schneider, D.: Death and cannibalism in a seasonal environment facilitate bacterial coexistence. Ecology letters, 12(1), 34-44 (2009).